Approaching Disparity Analysis

Region 5 Air Toxics Risk
Assessment Modeling Symposium
July 15, 2003

Alan Walts
Associate Regional Counsel

walts.alan@epa.gov

Purposes

- Enhance risk management
  - Environmental justice policy calls for "fair treatment" (no disproportionately high and adverse effects)
  - Focus on severity of risks alone may not support decisions in every case
  - Understanding risk distribution is part of characterizing risks
    - Usually thought of as population distribution; but can be geographic risk contours.
    - See EPA's Risk Characterization Handbook (Dec. 2000) (http://www.epa.gov/osp/spc/2riskchr.htm)

Purposes *(continued)*

- Investigate Title VI complaints (or ensure compliance before complaints are filed)
  - Background: see <http://www.epa.gov/civilrights/t6home.htm>
  - Role of risk assessment: determine whether alleged discriminatory impacts are "adverse"
  - Role of disparity analysis: determine whether "adverse" impacts are "disparate"
  - (Note: Discussions of disparity analysis here are illustrative, and don't necessarily describe how EPA will conduct disparity analysis in all cases under Title VI.)

Scoping the analysis

- Scoping is driven by the question posed, and by available ***data*** and ***tools***
- Need more EPA attention to developing best practices in disparity analysis
  - Examples of recent work using EPA air data:
    - Woodruff TJ, "Disparities in Exposure to Air Pollution During Pregnancy," *Environ Health Perspect* 111:942-946 (2003) (criteria pollutants).
    - Lopez R, "Segregation and Black/White Differences in Exposure to Air Toxics in 1990," *Environ Health Perspect* 110 (suppl 2):289-295 (2002) (using CEP data).
  - Approach described here may be easier to build into risk assessments

Data

- Risks
  - Dispersion modeling: the best means of describing extent and severity of exposure
- Population
  - Data from Census units
  - Use highest resolution available
    - Better to aggregate small units, than to estimate population location within large units

Data (cont.): describing spatial distribution of risks

- Two main options
  - Average impacts to census units (e.g., average cancer risk by census block)
  - Describe impacts as areas and assign population to areas
    - Receptor points → polygons
    - Receptor points → isopleths

Tools

- Visual correlation of risks and populations (using GIS)
- Statistical correlation of risks and populations

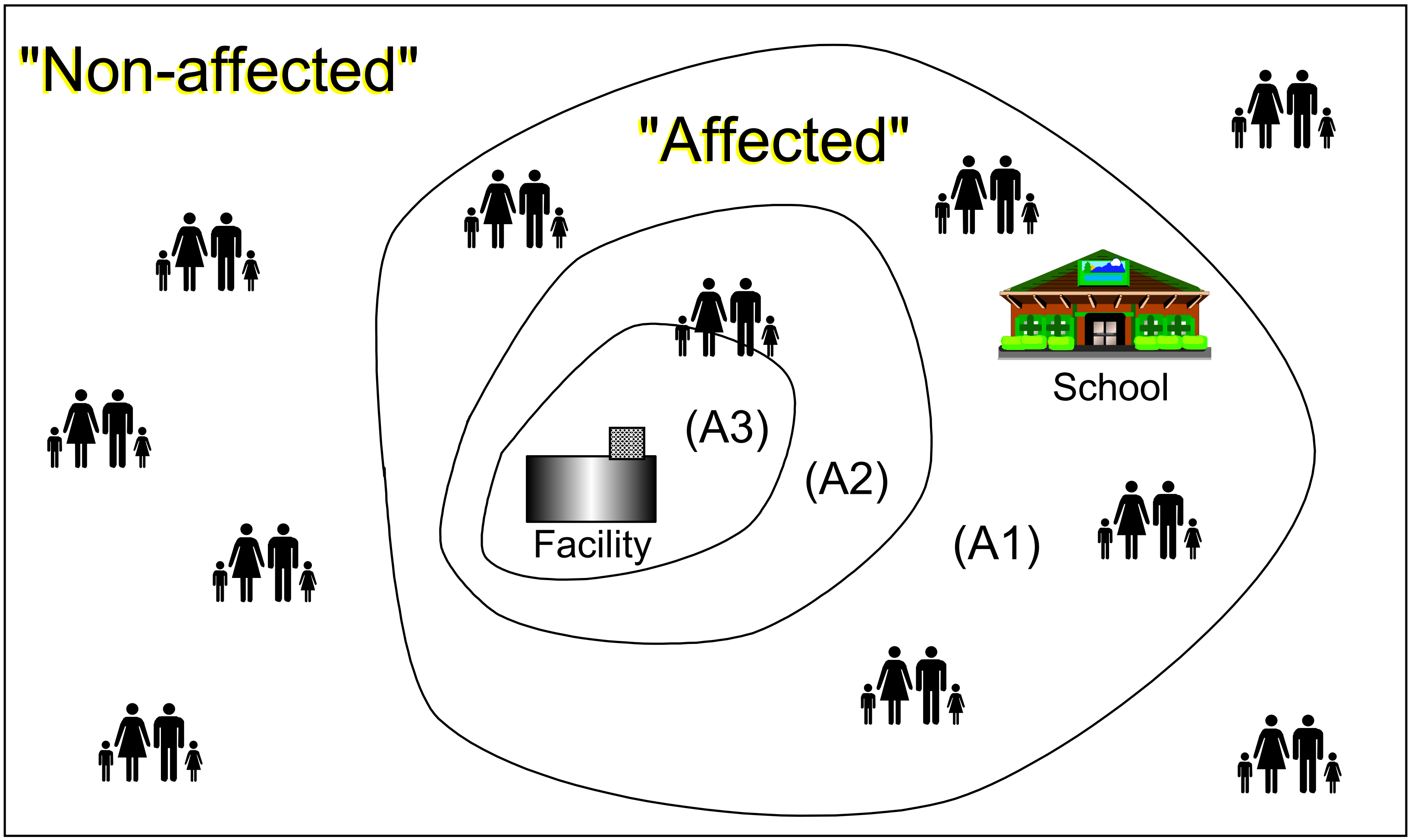
Tools: Visual correlation

- Good for broad screening, targeting, and description
- Doesn't offer clear & replicable decision criteria (numbers vs. "I know it when I see it")
- Not an either/or choice
- For recent review of some issues:

  Maantay J, "Mapping Environmental Injustices: Pitfalls and Potential of Geographic Information Systems in Assessing Environmental Health and Equity," *Environ Health Perpect* 110 (suppl 2): 161-171 (2002)

Example image:

risk isopleths in dispersion grid



Tools: Statistical correlation - first steps

- Identify "affected" and "comparison" populations
  - "Affected" based on relevant protectiveness standards
  - Select comparison pop. based on potential to be in affected pop. (helps keep variables constant)
  - Within each population, define group(s) based on parameter(s) of interest (Title VI: race, color, national origin; EJ/other: minority, low-income, children, subsistence farmers)

Tools: Defining "significance" of statistical correlation

- *Statistical* significance: the findings (based on a sample) are "real," not chance results
  - A common focus in caselaw on discriminatory impacts
- *Policy* significance: the "real" situation is worth doing something about

Statistical correlation - Relative ratio

- Used in Shintech Title VI investigation (see <http://www.epa.gov/civilrights/shinfileapr98.htm>)
- In Shintech, was applied to proximate vs. non-proximate populations (using circular buffer zones around TRI facilities).
- Could also be applied to dispersion modeling results (affected vs. non-affected populations).
  - The examples below assume dispersion modeling.

Example: data and hypotheses

	Group 1: Hispanic	Group 2: non-Hispanic	General population
Affected	800	2,000	2,800
Non-affected	200	2,000	2,200
Comparison	1,000	4,000	5,000

<u>Hypothesis</u>: Hispanics (Group 1) experience disparate impacts compared to non-Hispanics (Group 2). {ratio > 1}

<u>Null hypothesis</u>: Impacts on Hispanics (Group 1) are proportionate to their representation in the general population. {ratio = 1}

Relative ratio applied

	Group 1: Hispanic	Group 2: non-Hispanic	General population
Affected	800	2,000	2,800
Non-affected	200	2,000	2,200
Comparison	1,000	4,000	5,000

- p1 = (A Group 1 ÷ C Group 1) = (800 ÷ 1,000) = 0.8 x 100 = 80%
- p2 = (A Group 2 ÷ C Group 2) = (2,000 ÷ 4,000) = 0.5 x 100 = 50%

Relative ratio = p1 ÷ p2 = 80 ÷ 50 = 1.6

Or: Any given member of Group 1 is 60% more likely to be affected than any given member of Group 2

Statistical correlation - affected to non-affected

- Could also be applied to analyze one or more subsets of a given risk distribution (i.e., comparing more affected to less affected)
- Could supplement (or be used instead of) relative ratio analysis: a different way to define and measure disparity, using the same data.

Affected to non-affected applied

	Group 1: Hispanic	Group 2: non-Hispanic	General population
Affected	800	2,000	2,800
Non-affected	200	2,000	2,200
Comparison	1,000	4,000	5,000

- p1 = (A Group 1 ÷ A General pop.) = (800 ÷ 2,800) = 0.29 x 100 = 29%
- p2 = (non-A Group 1 ÷ non-A General pop.) = (200 ÷ 2,200) = 0.09 x 100 = 9%

Effects ratio = p1 ÷ p2 = 29 ÷ 9 = 3.22

<u>Or</u>: Any given member of Group 1 is 222% more likely to be affected than to be non-affected, compared to members of the general comparison population